applied sciences

MDPI

Article

Affective-Knowledge-Enhanced Graph Convolutional Networks for Aspect-Based Sentiment Analysis with Multi-Head Attention

Xiaodong Cui [1,2], Wenbiao Tao [1,2] and Xiaohui Cui [1,3,*]

[1] The Engineering Research Center of Cyberspace, Yunnan University, Kunming 650504, China; xdcui@mail.ynu.edu.cn (X.C.); taowenbiao@mail.ynu.edu.cn (W.T.)
[2] The Pilot School of Software, Yunnan University, Kunming 650504, China
[3] The School of Cyber Science Additionally, Enginning, Wuhan University, Wuhan 430072, China
* Correspondence: xcui@whu.edu.cn; Tel.: +86-133-0714-6750

Abstract: Aspect-based sentiment analysis (ABSA) is a task in natural language processing (NLP) that involves predicting the sentiment polarity towards a specific aspect in text. Graph neural networks (GNNs) have been shown to be effective tools for sentiment analysis tasks, but current research often overlooks affective information in the text, leading to irrelevant information being learned for specific aspects. To address this issue, we propose a novel GNN model, MHAKE-GCN, which is based on the graph convolutional neural network (GCN) and multi-head attention (MHA). Our model incorporates external sentiment knowledge into the GCN and fully extracts semantic and syntactic information from a sentence using MHA. By adding weights to sentiment words associated with aspect words, our model can better learn sentiment expressions related to specific aspects. Our model was evaluated on four publicly benchmark datasets and compared against twelve other methods. The results of the experiments demonstrate the effectiveness of the proposed model for the task of aspect-based sentiment analysis.

Keywords: graph convolutional network; aspect-based sentiment analysis; multi-head attention

check for updates

Citation: Cui, X.; Tao, W.; Cui, X. Affective-Knowledge-Enhanced Graph Convolutional Networks for Aspect-Based Sentiment Analysis with Multi-Head Attention. *Appl. Sci.* **2023**, *13*, 4458. https://doi.org/10.3390/app13074458

Academic Editor: Yu-Dong Zhang

Received: 28 February 2023
Revised: 22 March 2023
Accepted: 28 March 2023
Published: 31 March 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The field of sentiment analysis has gained significant attention in NLP due to the increasing popularity of mobile Internet and social media. One of the tasks within the field of sentiment analysis is ABSA, which focuses on identifying the sentiment (positive, negative, or neutral) associated with specific aspects within a sentence. For instance, in the review sentence "The food in the restaurant is delicious, but the service is poor," "food" and "service" are the two aspects, whereas "delicious" and "poor" are the corresponding affective words. As a result, the two aspects are associated with positive and negative sentiment polarities, respectively. However, a growing number of users are sharing their views online, but complex sentiment expressions make it especially difficult to capture what the users think. This makes some hard work for the ABSA task. The successful solution of the ABSA task allows the model to automatically identify fine-grained sentiment expressions, leading to excellent service and consumer experiences.

The original ABSA approach used feature words as the basic unit of analysis. For example, by analyzing which feature words in a text indicate sentiment and which sentiment polarity (positive, negative, neutral) these feature words correspond to. These features are usually words or phrases and can be obtained by manual definition and automatic extraction [1,2]. However, feature-based approaches face problems such as incomplete feature extraction and insufficient granularity. With the rise in sentiment lexicon, researchers have started to use sentiment lexicon to support ABSA, for instance, applying words from the sentiment lexicon to texts and associating them with aspects. This approach usually

requires consideration of factors such as the polarity and strength of sentiment words and the relationships between words. Over the past few years, the rapid progress in deep learning has made it possible for researchers to leverage neural networks in NLP tasks, including ABSA. Deep learning methods have significantly contributed to the development of ABSA techniques. ABSA methods based on neural networks often employ deep learning models such as convolutional neural networks (CNNs) [3] or recurrent neural networks (RNNs) [4] to capture the complex connections between aspects and emotions present in text data.

In ABSA, the syntactic structure can provide information about the association between a particular aspect and its corresponding sentiment expression. As a result, when predicting the sentiment polarity of a specific aspect, it is essential to consider the syntactic dependencies between contextual words and aspectual words [5,6]. These methods can extract long-term syntactic dependencies employing convolution.

Existing work that utilizes graph neural networks for sentiment analysis mostly focuses on syntactic dependency information while disregarding external affective knowledge related to the text. In our research, we demonstrate that incorporating external affective knowledge can improve the expression of a specific aspect. We achieve this by reconstructing the syntactic dependency tree and augmenting it with external affective knowledge from the sentiment lexicon. Specifically, we first use the original method to construct dependency graphs based on syntactic dependency trees. Then, we incorporate external affective knowledge by assigning each word in the sentence a sentiment score from the sentiment lexicon during the graph construction process. As a result, each sentence can be modeled as a dependency graph that captures both the dependency relationships between the context and the specific aspect. The resulting affective-knowledge-enhanced dependency graph is then fed into a graph convolutional neural network-based model. Furthermore, the conventional attention mechanism can introduce a considerable amount of noise because of its widely scattered weight values. As a result, it can severely impact the effectiveness of the model. We utilize multi-head attention to tackle this issue, as it can effectively overcome the limitations of traditional attention and improve the model's overall performance. Multi-head attention introduces multiple subspaces and heads, which work in tandem to learn and attend to the different parts of the input. This approach allows the model to capture a wider range of nuanced and varied information from the text, resulting in improved performance for the ABSA task. The main contributions of this paper are as follows:

- Considering the influences of sentiment words carried by the sentences themselves on sentiment analysis, an external sentiment lexicon was introduced to reconstruct the graph based on the dependency tree, increasing the weights of sentiment words so that the network model could strengthen the dependency between sentiment words and aspects.
- Since syntactic relations help to enhance sentiment analysis tasks, the syntactic relations of sentences are introduced into the network model by means of graph convolutional networks.
- The multi-head self-attention mechanism extracts both global semantic information from the context and local semantic information from the aspect. The graph convolutional network extracts syntactic relations, which are then fused with the global and local semantics using multi-head interactive attention. This enables the network model to learn both rich semantic and syntactic relational dependencies.

2. Related Work

We categorize previous related work into three distinct parts for the sake of brevity: aspect-based sentiment analysis methods, graph convolutional neural network methods, and affective-knowledge-enhanced methods.

2.1. Aspect-Based Sentiment Analysis

Sentiment analysis is a task that aims to determine the sentiment polarity of a given text. ABSA is a subtask of sentiment analysis that specifically analyzes the sentiment polarity of specific aspects. Tang et al. [7] used two long short-term memory (LSTM) networks to model the text on both sides of a specific aspect, enabling the extraction of bidirectional sentiment information. This approach captures the sentiment information associated with a specific aspect of a sentence. Ma et al. [8] proposed the interactive attention network (IAN) model, which utilizes two attention networks to model the interaction between the aspect and context. This allows the model to concentrate on the important parts of both the aspect and context, generating the representations of the aspect and context. By effectively learning features of both the aspect and context, the method provides ample information to ascertain the sentiment polarity of the aspect. In their work, Xue and Li [3] introduced a gated convolutional neural network model that employs gated units to selectively generate sentiment features pertaining to a specific aspect. Huang et al. [9] used an attention-over-attention (AOA) model to focus on key parts of sentences and learned the interaction between the context and the aspect. While these models show some performance advantages with the help of attention and syntactic information, they ignore the dependencies between the context and aspect, which are crucial for the ABSA task.

2.2. Graph Convolutional Network

The area of NLP has seen a surge in interest towards GNN-based models in recent years. This is largely due to their impressive achievements in multiple NLP domains, including text classification [10,11], relationship extraction [12], named entity recognition (NER) [13], and more. In ABSA, graph neural networks have shown great potential. As an example, Zhang et al. [14] were pioneers in utilizing GCN for ABSA. They suggested the implementation of GCN on syntactic dependency trees to effectively leverage contextual and aspectual dependency information within sentences. Other studies, such as Huang et al. [15] and Sun et al. [16], have also proposed graph neural network-based models for ABSA. With evermore graph neural network-based models demonstrating their superior performance, it has been shown that they are very effective at strengthening the contextual and aspectual dependencies in ABSA tasks with the help of graph neural networks. However, these models overlook the sentiment information in contextual and aspectual words; thus, the incorporation of affective knowledge to enhance the feature-representation extraction by the GCN in the ABSA task needs to be considered.

2.3. Affective Knowledge

External knowledge is a crucial factor in numerous NLP tasks. [17,18]. In sentiment analysis tasks, external sentiment knowledge is also commonly used as a source for enhancing the representation of sentiment features [19–21]. SenticNet is an open-source resource for sentiment analysis that infers the sentiment polarity of commonsense concepts through a dimensionality reduction approach and assigns an affective value to each concept [22–27]. A concept with an affective value approaching one suggests a more positive sentiment, whereas an affective value nearing negative one suggests a more negative sentiment. On the other hand, an affective value of zero indicates a neutral sentiment. In the ABSA task, with the help of SenticNet, a sentiment lexicon, the expressions of sentiment words with dependencies on aspects in the context can be fully explored. Liang et al. [28] utilized SenticNet to inject affective knowledge based on the model proposed by Zhang et al. [14] and obtained a significant performance improvement on benchmark datasets. The results show that the inclusion of external affective knowledge in the ABSA task is effective at improving the capturing of aspect-specific affective features.

3. Methodology

This section presents our model (MHAKE-GCN), which is depicted in Figure 1 and is explained in detail below.

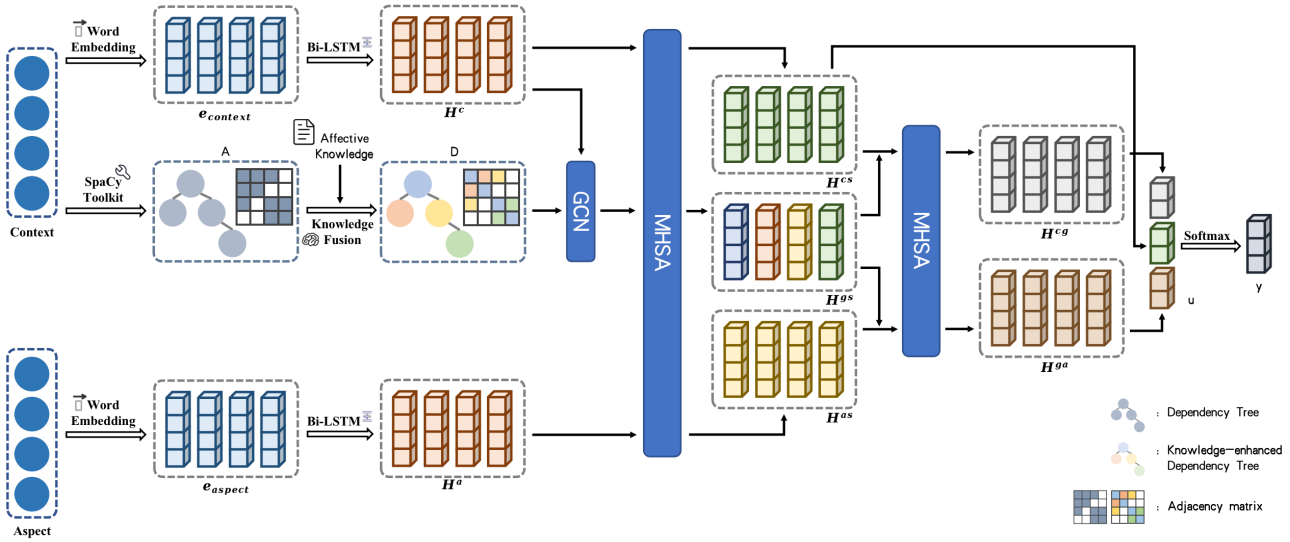


Figure 1. The overall structure of the proposed MHAKN-GCN. The characters in the above figure such as A, D, u, y, etc. correspond to A, D, u, y, etc. in the equation below.

3.1. Task Definition

Assume S is a sentence composed of n contextual words and m aspectual words. Let us denote the context and aspect separately as $[w_1^c, w_2^c, \cdots, w_n^c]$ and $[w_1^a, w_2^a, \cdots, w_m^a]$, respectively. Here, w_i^c indicates the ith contextual word, and w_j^a indicates the jth aspectual word. It is crucial to recognize that a sentence may include multiple aspects, each corresponding to a different sentiment polarity, such as positive, neutral, or negative. The main objective of ABSA is to identify the sentiment polarity of a given aspect by examining the aspect-related affective information present in the context.

3.2. Embedding Module

To obtain embeddings for each word in the input context and aspect, we initially utilize an embedding lookup table denoted as $E \in R^{k \times |N|}$, where k represents the dimension of the word vector and $|N|$ is the size of the vocabulary. This lookup table maps every word to a k-dimensional embedding. This results in the context embedding matrix $e_{context} = [e_1^c, e_2^c, \cdots, e_n^c]$ and aspect embedding matrix $e_{aspect} = [e_1^a, e_2^a, \cdots, e_m^a]$, where $e_i^c \in R^k$ is the embedding of the i-th word in the context w_i^c, and $e_j^a \in R^k$ is the embedding of the j-th word in the aspect w_j^a. These embeddings can be obtained from pre-trained models such as GloVe [29] or BERT [30].

3.3. Constructing Graph

Following the approach proposed in [14,16], we construct the graph convolutional network for each input sentence over the dependency tree. The adjacency matrix $A \in R^{n \times n}$ of the sentence is obtained as follows:

$$A_{i,j} = \begin{cases} 1, & w_i \ \ and \ \ w_j \ \ have \ \ dependences \\ 0, & otherwise \end{cases} \quad (1)$$

According to the former GCN-based model [14], we construct an undirected graph, with parent and child nodes influencing each other.

3.4. Reconstructing Graph with Enhanced Affective Knowledge

To effectively utilize the affective information between the context and aspect, we enhance the representation of the adjacency matrix by leveraging the affective score from SenticNet [31]:

$$S_{i,j} = SenticNet(w_i) + SenticNet(w_j) \tag{2}$$

where $SenticNet(w_i)$ is the affective score from the SenticNet sentiment lexicon. It is worth noting that if the affective score of w_i is 0, then w_i is neutral or not in the SenticNet sentiment lexicon. We can then extract the affective dependencies from the two dependent nodes, and the sentiment information of the two dependent nodes depends on the sum of their affective scores. Samples of concepts and their corresponding affective scores are listed in Table 1.

Table 1. Samples of affective words in SenticNet.

Concept	Intensity
smile	0.997
wonderful	0.805
everything	0.099
hour	0.093
sad	−0.910
uncomfortable	−0.900

Further, to enhance the dependency relationship between the context and the aspect, consider whether the two nodes where the dependency exists contain aspectual words:

$$T_{i,j} = \begin{cases} 1, & w_i \ \ or \ \ w_j \ \ is \ \ a \ \ aspect \ \ word \\ 0, & otherwise \end{cases} \tag{3}$$

Next, we can get the affective-knowledge-enhanced adjacency matrix $D_{i,j}$ of the sentence:

$$D_{i,j} = A_{i,j} \times (S_{i,j} + T_{i,j} + 1) \tag{4}$$

Algorithm 1 outlines the procedure for generating the affective adjacency matrix for each sentence.

Algorithm 1 The generation of the affective adjacency matrix for each sentence.

Require: a sentence $W^c = \{w_1^c, w_2^c, \cdots, w_n^c\}$; a aspect $W^a = \{w_1^a, w_2^a, \cdots, w_m^a\}$; the dependency tree of the sentence $dependency(W^c)$; a collection of sentiment words generated by SenticNet.

for $i = 1 \rightarrow n$ **do**
  for $j = 1 \rightarrow n$ **do**
    if $dependency(w_i^c, w_j^c) \in dependency(dependency(W^c))$ **or** $i = j$ **then**
      $A_{i,j} \leftarrow 1$ ▷ Generated by dependency tree
      $S_{i,j} \leftarrow SenticNet(w_i^c) + SenticNet(w_j^c)$ ▷ Enhanced by SenticNet
      if $w_i^c \in W^a$ **or** $w_j^c \in W^a$ **then**
        $T_{i,j} \leftarrow 1$
      else
        $T_{i,j} \leftarrow 0$
      end if
      $D_{i,j} \leftarrow A_{i,j} \times (S_{i,j} + T_{i,j} + 1)$
    else
      $D_{i,j} \leftarrow 0$
    end if
  end for
end for

3.5. Graph Convolutional Networks

To apply a GCN based on a dependency tree [14], we begin by constructing a primitive dependency tree X for a given sentence S consisting of k words. The tree has k nodes, one for each word, with edges representing dependencies between words. We represent the original dependency matrix as $A_{ij} \in R^{k \times k}$, where a value of 1 indicates the existence of a dependency between nodes and 0 indicates no dependency. For the i-th node, we denote the hidden state representation evolved from previous GCN layers as h_i^{l-1} and the output as h_i^l. The original GCN is then computed as follows:

$$h_i^l = \sigma(\sum_{j=1}^{n} A_{ij} W^l h_i^{l-1} + b^l) \tag{5}$$

The trainable parameters of the model include the weights W^l and bias b^l, and σ represents a non-linear function, such as ReLU. Then, we replace the original dependency matrix A_{ij} with the knowledge-enhanced dependency matrix W_{ij}. We receive the ultimate representation of the hidden state H after GCN:

$$H = \sigma(\sum_{j=1}^{n} D_{ij} W^l h_i^{l-1} + b^l) \tag{6}$$

3.6. Multi-Head Attention

MHA [32] enables the creation of various projection information in multiple projection spaces. In our model, we utilize multi-head self-attention and multi-head interactive attention to model different aspects. Initially, we define a key sequence $k = k_1, k_2, \cdots, k_n$ and a query sequence $q = q_1, q_2, \cdots, q_n$. Next, the attention distribution is calculated based on the key and then attached to the value to compute the attention score. Typically, in NLP applications, keys and values are set to be equal—i.e., key = value. Finally, the key and query are projected to an output sequence using an attention function:

$$Attention(k, q) = softmax(f_m(k, q))k \tag{7}$$

f_m is the function utilized to compute the semantic correlation between k_i and q_j:

$$f_m(k_i, q_j) = tanh([k_i; q_j] \cdot W_a) \tag{8}$$

where W_a is the learnable weight matrix. MHA can learn *n_head* different scores in parallel subspaces. The outputs of *n_head* are concatenated as follows:

$$MHA(k, q) = [o^1 \oplus o^2 \oplus \cdots \oplus o^{n_{head}}] \cdot W_m \tag{9}$$

$$o^i = Attention^i(k, q) \tag{10}$$

The concatenation of vectors is represented by the symbol "$\oplus$", and W_m is a trainable weight matrix. The attention function generates the output o^i of the ith head. MHA can be the situation of MHSA when the query and key sequences are the same, $q = k$. MHAKE-GCN utilizes semantic encoding to obtain the hidden state of the context H^{cs} and the hidden state of the aspect H^{as} by leveraging the hidden states of the context H^c and the hidden states of the aspect H^a:

$$H^{cs} = MHA(H^c, H^c) \tag{11}$$

$$H^{as} = MHA(H^a, H^a) \tag{12}$$

After passing through L layers of GCN, the ultimate output is $H^L = h_1^L, h_2^L, \cdots, h_n^L$. We obtain the hidden state H^{gs} after semantic encoding by multi-head self-attention from the output of the last GCN layer as follows:

$$H^{gs} = MHA(H^L, H^L) \tag{13}$$

Multi-head interactive attention (MHIA) is a variant of MHA in which the query q is different from the key k. the context-aware syntactical information H^{cg} and the syntax-aware aspect information H^{ga} are obtained by:

$$H^{cg} = MHA(H^{cs}, H^{gs}) \tag{14}$$

$$H^{ga} = MHA(H^{gs}, H^{as}) \tag{15}$$

3.7. Information Fusion

To obtain the final feature representation u, we use average pooling to average the context-aware syntactical representation H^{cg}, the syntax-aware aspect H^{ga}, and the context semantic encoding H^{cs}, before concatenating them together:

$$h_{avg}^{cg} = \sum_{i=1}^{n} h_i^{cg}/n \tag{16}$$

$$h_{avg}^{ga} = \sum_{i=1}^{n} h_i^{ga}/n \tag{17}$$

$$h_{avg}^{cs} = \sum_{i=1}^{n} h_i^{cs}/n \tag{18}$$

$$u = [h_{avg}^{cg} \oplus h_{avg}^{cs} \oplus h_{avg}^{ga}] \tag{19}$$

3.8. Sentiment Classification

We feed the final obtained feature representation u into the $softmax$ layer, which allows us to obtain the probability distribution for the different aspects of sentiment polarity:

$$x = W_u^T u + b_u \tag{20}$$

$$y = softmax(x) \tag{21}$$

3.9. Model Training

The parameters of our model are updated by the gradient descent algorithm. The goal of training the model is to minimize the cross-entropy loss with L2 regularization:

$$L = -\sum_{i=1}^{S}\sum_{j=1}^{C} \hat{y}_i^j log y_i^j + \lambda ||\Theta||^2 \tag{22}$$

where S is the number of training samples and C is the number of sentiment categories. $\hat{y}$ is the correct distribution of sentiment. Θ represents all trainable parameters. λ is the parameter of regularization.

4. Experiments

4.1. Datasets

We evaluated the performance of our model by conducting experiments on four public datasets (Restaurant14, Laptop14, Restaurant15, and Restaurant16). The Restaurant14 and

Laptop14 datasets are from the semeval-2014 task [33], the Restaurant15 dataset is from the semeval-2015 task [34], and the Restaurant16 dataset is from the semeval-2016 task [35]. Each sample in the datasets comprises a comment sentence, an aspect containing one or more words, and the sentiment polarity corresponding to the aspect. Table 2 presents the distribution of sentiments in the four datasets.

Table 2. Statistics on the distribution of sentiments in the four datasets.

Dataset	Positive		Neural		Negative	
	Train	Test	Train	Test	Train	Test
Restaurant14	2164	728	637	196	807	196
Laptop14	996	341	464	169	870	128
Restaurant15	1178	439	50	35	382	328
Restaurant16	1620	597	88	38	709	190

4.2. Experimental Parameter Settings

In our experiments, we used the non-BERT-based model MHAKE-GCN, which involves embedding each word into a 300-dimensional GloVe vector [29]. The GCN consists of two layers, with a coefficient λ of 0.00001 for L_2 regularization. We used Adam [36] for parameter initialization and updates, with a learning rate of 0.001 and a batch size of 32. For the BERT-based model, MHAKE-GCN-BERT, the word embedding dimension was set to 768. The learning rate was 0.00002, and the batch size was 32.

4.3. Baseline Model

We introduce twelve outstanding models for comparison with our model to validate the performance of our model. We simply divide these models into four categories: semantic and attention-based models, GNN-based models, BERT-based models, and our model. The models we used follow:

- **SVM** [37] is a support vector machine method that achieves classification and regression by finding the optimal hyperplane.
- **GCAE** [3] selectively outputs sentiment features of the specific aspects with a gated CNN model.
- **IAN** [8] concentrates on interactive modeling of the aspect and the context.
- **AOA** [9] uses an attention-over-attention model to focus on key parts of sentences and learns the interaction between the context and the aspect.
- **ASGCN-DT** [14] gets a directional graph as input to a GCN, extracting the syntactic and dependency information of the context.
- **ASGCN-DG** [14] is the same as ASGCN-DT, with the only difference being that the adjacency matrix of the graph in ASGCN-DG is undirected.
- **CDT** [16] obtains the contextual dependencies related to the aspect by applying a GCN to the dependency tree generated from the sentence.
- **R-GAT** [38] explores a relational graph attention network for reconstructing an aspect-based tree structure, which is then used for sentiment classification.
- **AEN-BERT** [32] uses an attentional encoder network to model the context and target.
- **BERT-SPC** [32] is a varient of the transformer.
- **TD-BERT** [39] is a BERT-based model for target-dependent sentiment analysis.
- **R-GAT-BERT** [38] is the same as R-GAT, except replacing Bi-LSTM with BERT.
- **MHAKE-GCN** is the model we proposed.
- **MHAKE-GCN-BERT** is the same as MHAKE-GCN. The only difference is that the Bi-LSTM is substituted with BERT.

4.4. Experimental Results

The experimental results in Table 3 demonstrate that our model, MHAKE-GCN, outperformed the other compared models on all four benchmark datasets. As for the

syntactic and attention-based approaches, the GNN-based model outperformed them across the board according to two evaluation metrics on the four datasets. The best graph neural network-based approach, R-GAT, outperformed the best syntactic and attention-based approach, AOA, by 3.42%, 1.12%, 2.13%, and 1.42% in accuracy and 6.91%, 8.04%, 7.15%, and 4.68% in F1 for the four datasets, respectively. When compared with syntactic and attention-based approaches, MHAKE-GCN significantly outperformed them on all four datasets, indicating that our model, which uses a GCN based on a dependency tree, is better at capturing the dependencies between the context and the aspects. Compared with the graph neural network-based approach, MHAKE-GCN also showed superior performance in most metrics. There were only slight performance differences on the F1 metric for Restaurant14 and Restaurant16 datasets compared to the R-GAT model, suggesting that our model, which incorporates external sentiment knowledge and uses multi-head attention to fuse sentiment information, is better at extracting aspect-specific sentiment information, whereas R-GAT can construct aspect-related dependencies well due to its use of aspects as root nodes. Our model, which is constructed using traditional dependencies, is 2.4% and 1.57% behind R-GAT in terms of F1 score for Restaurant14 and Restaurant16, respectively. In addition, compared with the best BERT-based results, MHAKE-GCN-BERT, our BERT-based model achieved accuracies 1.05%, 1.37%, 0.78%, and 2.19% higher, and the F1 metrics were 1.20%, 2.88%, 1.21%, and 2.38% higher for the four datasets, respectively. Our proposed model outperformed the other BERT-based models used in the comparison, indicating that our model can achieve comparative performance even when encoded using a powerful pre-trained language model.

Table 3. Experimental results on four datasets. The best results are shown in bold.

Models		**Restaurant14 (%)**		**Laptop14 (%)**		**Restaurant15 (%)**		**Restaurant16 (%)**	
		Accuracy	**F1**	**Accuracy**	**F1**	**Accuracy**	**F1**	**Accuracy**	**F1**
Syn. + Att	SVM	80.16	-	70.49	-	-	-	-	-
	GCAE	77.28	-	69.14	-	-	-	-	-
	IAN	79.26	70.09	72.05	67.38	78.54	52.65	84.74	55.21
	AOA	79.97	70.42	72.62	67.52	78.17	57.02	87.50	66.21
Graph	ASGCN-DT	80.86	72.19	74.14	69.24	79.34	60.78	88.69	66.64
	ASGCN-DG	80.77	72.02	75.55	71.05	79.89	61.89	88.99	67.48
	MAGCN	81.25	71.94	75.39	72.47	-	-	-	-
	R-GAT	83.39	**77.33**	73.74	75.56	80.30	64.17	88.92	**70.89**
Ours	MHAKE-GCN	**83.96**	74.93	**78.46**	**76.54**	**82.10**	**65.90**	**89.36**	69.32
BERT	AEN-BERT	83.12	73.76	79.93	76.31	-	-	-	-
	BERT-SPC	84.46	76.98	78.99	75.03	-	-	-	-
	TD-BERT	85.10	78.35	78.87	74.38	-	-	-	-
	R-GAT-BERT	86.60	81.35	78.21	74.07	83.22	69.79	89.71	76.62
Ours	MHAKE-GCN-BERT	**87.65**	**82.55**	**81.30**	**79.19**	**84.00**	**71.00**	**91.90**	**79.00**

4.5. Ablation Experiment

In order to further validate the effectiveness of each component in our model MHAKE-GCN, we conducted ablation experiments, and the results are presented in Table 4. The "w/o" in the table denotes "without".

Table 4. The ablation experiment of our MHAKE-GCN.

Models	Restaurant14 (%)		Laptop14 (%)		Restaurant15 (%)		Restaurant16 (%)	
	Acc.	**F1**	**Acc.**	**F1**	**Acc.**	**F1**	**Acc.**	**F1**
MHAKE-GCN w/o MHSA	82.16	74.28	77.64	75.44	80.29	64.28	87.64	68.87
MHAKE-GCN w/o MHIA	81.59	72.47	74.21	71.80	79.03	61.94	86.83	68.96
MHAKE-GCN w/o Affective Knowledge	81.04	71.36	75.91	71.63	79.95	60.87	87.38	68.21
MHAKE-GCN	**83.96**	**74.93**	**78.46**	**76.54**	**82.10**	**65.90**	**89.36**	**69.32**

The results of the ablation experiments indicate that all components in our model are important, and their absence leads to decreased performance. The MHSA component extracts rich semantic information and encodes syntactic information. Ablating MHSA resulted in slight decreases in metrics on all four datasets. Of these, the F1 metrics for the Laptop14 and Restaurant15 datasets decreased a bit more, 1.10% and 1.62% respectively, indicating that syntactic information is useful for ABSA. The MHIA component aggregates features and learns interactive associations between semantic and syntactic information. Ablating MHIA resulted in significant decreases in metrics on all four datasets. The Laptop14 and Restaurant15 datasets showed particularly severe performance degradation, with the accuracy and F1 metrics dropping by 4.25% and 4.74%, respectively, for the Laptop14 dataset, and the F1 metric dropping by 4.04% for the Restaurant15 dataset, indicating that the interaction between semantic and syntactic information is crucial for ABSA. Finally, ablation of the affective knowledge component resulted in significant performance reduction on all four datasets. The F1 metric decreased by 3.57% for the Restaurant14 dataset, 4.91% for the Laptop14 dataset, and 5.03% for the Restaurant15 dataset, demonstrating that incorporating external affective knowledge can significantly improve ABSA performance.

4.6. Effect of a GCN Layer Number

In this section, we examine how the number of GCN layers in our model, MHAKE-GCN, affects its performance on four datasets. We conducted experiments by setting the number of GCN layers to 1 to 8 and evaluated the model's accuracy and macro-F1 scores. The experimental results are presented in Figures 2 and 3.

The figures show that both the accuracy and macro-F1 scores reach their highest values when the number of GCN layers is two. When the number of GCN layers is one, the model does not adequately learn the sentiment dependence of specific aspects, resulting in poor performance. On the other hand, when the number of GCN layers exceeds two, the performance drops as the number of layers increases. This suggests that increasing the model's parameters sharply reduces its efficiency.

4.7. Effect of Training Sample Size

We investigated whether the performance of the model is affected by the number of training samples and conducted an experiment by randomly extracting a certain percentage of samples from the original training set. The experimental results are depicted in Figures 4 and 5.

As depicted in the above graphs, the accuracy and macro-F1 score reached their maxima when the number of training samples was seventy percent. When the number of training samples was too small, the model was not sufficiently trained, resulting in an under-fitted model. Conversely, when there were too many training samples, the model learned too much irrelevant information, causing a degradation in its performance.

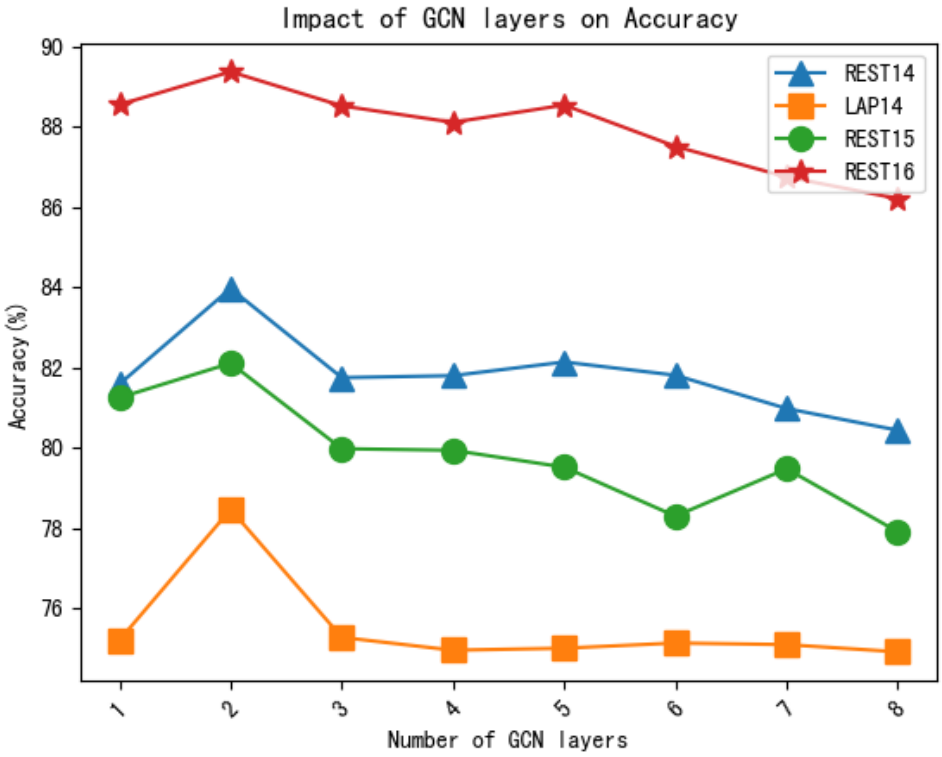


Figure 2. Effect of GCN layers on accuracy.

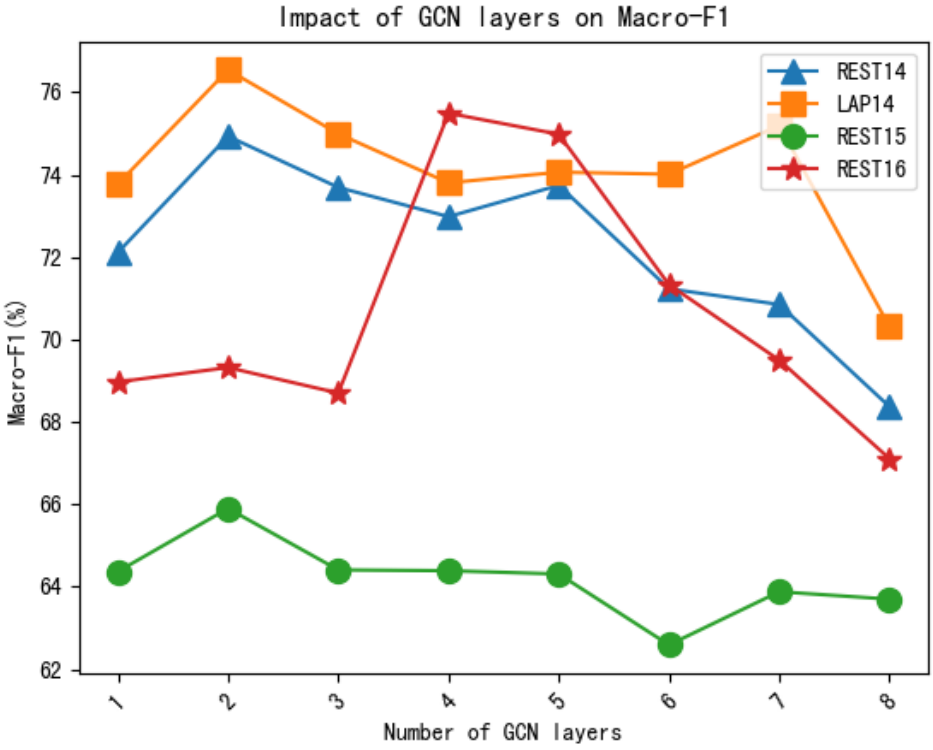


Figure 3. Effect of GCN layers on macro-F1.

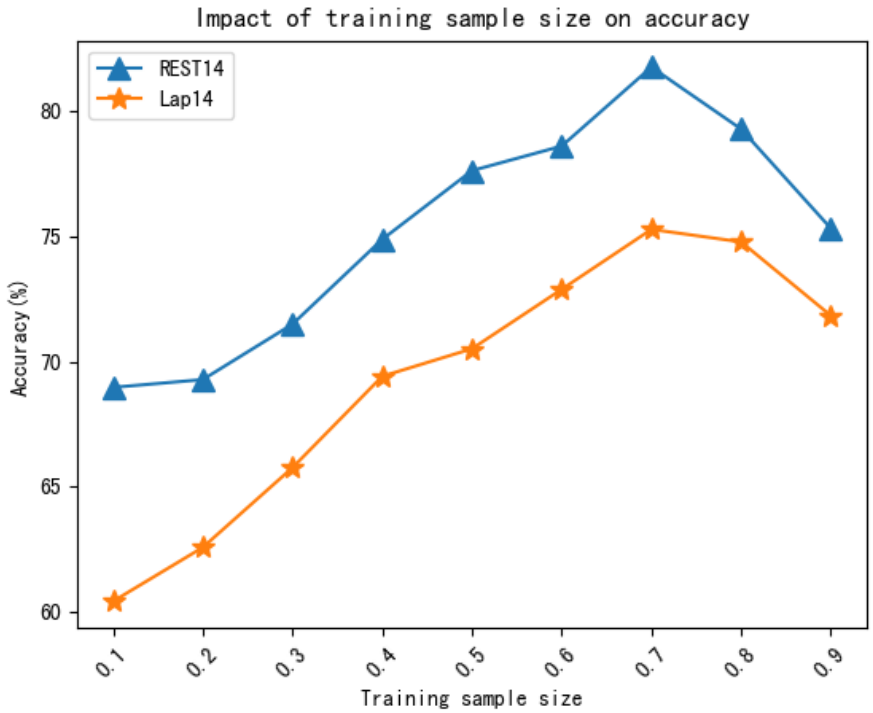


Figure 4. Effect of training sample size on accuracy.

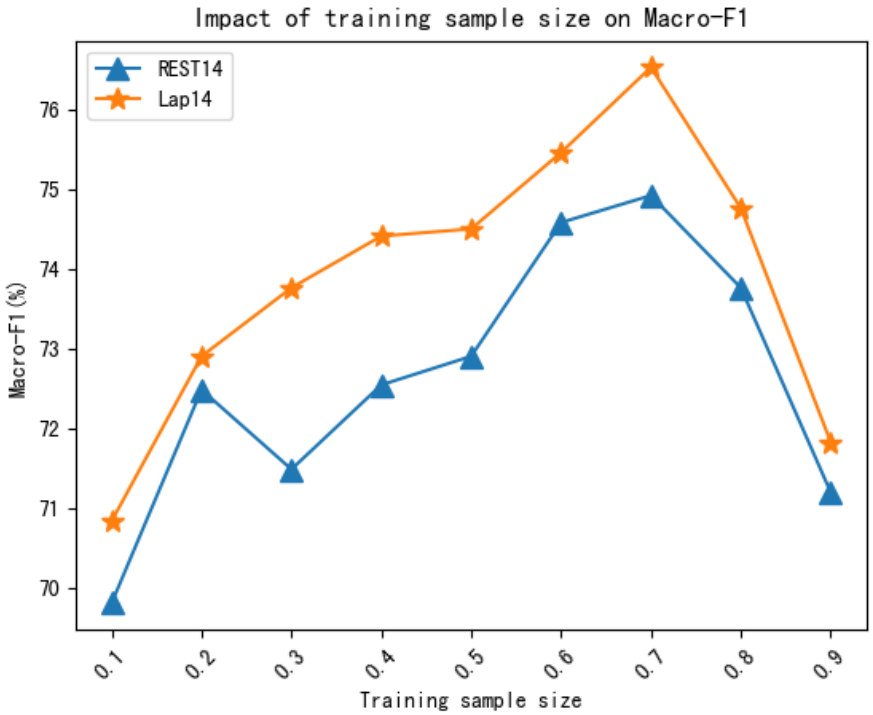


Figure 5. Effect of training sample size on macro-F1.

5. Discussion

The model we propose in this paper, MHAKE-GCN, was verified to perform competitively by the above comparison experiments and ablation experiments. Previous syntactic and attention-based models [3,8,9,37] have considered the need to exploit both syntactic and semantic information in context, but previous models ignored the associations of context with specific aspects, resulting in models that introduce too many aspect-irrelevant sentiment representations, making performance unsatisfactory. To compensate for the

shortcoming mentioned above, models based on graph neural networks [14,16,38] have emerged that construct graph representations based on syntactic dependency trees and then extract syntactic information through graph neural networks, which can adequately extract aspect-specific sentiment information. In the above graph neural network-based models, words with dependency relationships with aspects are assigned the same weight during the construction of the dependency graph; however, there are often many words denoting a sentiment in the context, and these sentiment words tend to influence the sentiment tendency of a particular aspect. The results of the comparison experiment and the results of the ablation experiment demonstrated the effectiveness of incorporating external affective knowledge. In the ablation experiment, it could be seen that multi-head attention in the model has a positive effect on the ABSA task, fully extracting syntactic and semantic information from the context and the aspect.

6. Implications

The model we propose can be well applied to real-life situations where the Internet allows people to express their views and comments. For example, if a customer gives a positive opinion about the food in a restaurant and a negative opinion about the service, the business owner can use our model to extract the positive sentiment corresponding to the food aspect and the negative sentiment corresponding to the service aspect, so as to improve the quality of the service and thereby improve the customer experience.

Our model can be applied to a variety of situations, such as movie reviews, product purchase reviews, social media discussions, and so on. However, since the corpus we trained on has information about restaurants and laptops, applications in these two domains will be better.

7. Conclusions and Future

In this paper, a model combining affective-knowledge-enhanced graph convolutional networks and multi-head attention was presented. In order to overcome the problem of long-distance sentiment loss, Bi-LSTM was used for encoding, followed by extracting the respective semantic information of the context and aspects via MHIA. For the purpose of making full use of syntactic information and considering the influences of sentiment words in the context, we proposed affective-knowledge-enhanced graph convolutional networks. The semantic interaction is then performed via MHIA. Experimental results on four benchmark datasets show that our model, MHAKE-GCN, outperforms traditional syntactic and attention-based models and graph neural network-based models, proving its effectiveness.

Although existing models have achieved excellent results, there is still much to explore in ABSA. In the future, we intend to try to reduce the parameters to make the model more lightweight or to explore the potential expression of emotions.

Author Contributions: Conceptualization, X.C. (Xiaodong Cui) and X.C. (Xiaohui Cui); methodology, X.C. (Xiaodong Cui); validation, X.C. (Xiaodong Cui); investigation, X.C. (Xiaodong Cui); resources, X.C. (Xiaohui Cui); writing—original draft preparation, X.C. (Xiaodong Cui); writing—review and editing, X.C. (Xiaodong Cui) and W.T.; visualization, X.C. (Xiaodong Cui); project administration, X.C. (Xiaodong Cui); funding acquisition, X.C. (Xiaohui Cui). All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Yunnan Province Science Foundation under Grant No. 202001BB050076, in part by the Fund Project of Yunnan Province Education Department under Grant No. 2022j0008.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Popescu, A.M.; Etzioni, O. Extracting product features and opinions from reviews. In *Natural Language Processing and Text Mining*; Springer: Cham, Switzerland, 2007; pp. 9–28.
2. Hu, M.; Liu, B. Mining and summarizing customer reviews. In Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA, USA, 22–25 August 2004; pp. 168–177.
3. Xue, W.; Li, T. Aspect based sentiment analysis with gated convolutional networks. *arXiv* **2018**, arXiv:1805.07043.
4. Mikolov, T.; Karafiát, M.; Burget, L.; Cernockỳ, J.; Khudanpur, S. Recurrent neural network based language model. In Proceedings of the Interspeech, Chiba, Japan, 26–30 September 2010; Volume 2, pp. 1045–1048.
5. Yang, P.; Li, L.; Luo, F.; Liu, T.; Sun, X. Enhancing topic-to-essay generation with external commonsense knowledge. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 2002–2012.
6. Parthasarathi, P.; Pineau, J. Extending neural generative conversational model using external knowledge sources. *arXiv* **2018**, arXiv:1809.05524.
7. Tang, D.; Qin, B.; Feng, X.; Liu, T. Effective LSTMs for target-dependent sentiment classification. *arXiv* **2015**, arXiv:1512.01100.
8. Ma, D.; Li, S.; Zhang, X.; Wang, H. Interactive attention networks for aspect-level sentiment classification. *arXiv* **2017**, arXiv:1709.00893.
9. Huang, B.; Ou, Y.; Carley, K.M. Aspect level sentiment classification with attention-over-attention neural networks. In Proceedings of the Social, Cultural, and Behavioral Modeling: 11th International Conference, SBP-BRiMS 2018, Washington, DC, USA, 10–13 July 2018; Springer: Cham, Switzerland, 2018; pp. 197–206.
10. Piao, Y.; Lee, S.; Lee, D.; Kim, S. Sparse Structure Learning via Graph Neural Networks for Inductive Document Classification. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2022; Volume 36; pp. 11165–11173.
11. Yao, L.; Mao, C.; Luo, Y. Graph convolutional networks for text classification. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 7370–7377.
12. Tian, Y.; Chen, G.; Song, Y.; Wan, X. Dependency-driven relation extraction with attentive graph convolutional networks. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*; Association for Computational Linguistics: Cedarville, OH, USA, 2021; pp. 4458–4471.
13. Ding, R.; Xie, P.; Zhang, X.; Lu, W.; Li, L.; Si, L. A neural multi-digraph model for Chinese NER with gazetteers. In Proceedings of the 7th Annual Meeting of the Association for Computational Linguistics, Beijing, China, 26–31 July 2019; pp. 1462–1467.
14. Zhang, C.; Li, Q.; Song, D. Aspect-based sentiment classification with aspect-specific graph convolutional networks. *arXiv* **2019**, arXiv:1909.03477.
15. Huang, B.; Carley, K.M. Syntax-aware aspect level sentiment classification with graph attention networks. *arXiv* **2019**, arXiv:1909.02606.
16. Sun, K.; Zhang, R.; Mensah, S.; Mao, Y.; Liu, X. Aspect-level sentiment analysis via convolution over dependency tree. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 5679–5688.
17. Young, T.; Cambria, E.; Chaturvedi, I.; Zhou, H.; Biswas, S.; Huang, M. Augmenting end-to-end dialogue systems with commonsense knowledge. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.
18. Chaturvedi, I.; Satapathy, R.; Cavallari, S.; Cambria, E. Fuzzy commonsense reasoning for multimodal sentiment analysis. *Pattern Recognit. Lett.* **2019**, *125*, 264–270. [CrossRef]
19. Poria, S.; Chaturvedi, I.; Cambria, E.; Bisio, F. Sentic LDA: Improving on LDA with semantic similarity for aspect-based sentiment analysis. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 4465–4473.
20. Dragoni, M.; Poria, S.; Cambria, E. OntoSenticNet: A commonsense ontology for sentiment analysis. *IEEE Intell. Syst.* **2018**, *33*, 77–85. [CrossRef]
21. Dragoni, M.; Donadello, I.; Cambria, E. OntoSenticNet 2: Enhancing reasoning within sentiment analysis. *IEEE Intell. Syst.* **2022**, *37*, 103–110. [CrossRef]
22. Cambria, E.; Speer, R.; Havasi, C.; Hussain, A. Senticnet: A publicly available semantic resource for opinion mining. In Proceedings of the AAAI Fall Symposium: Commonsense Knowledge, Arlington, VA, USA, 11–13 November 2010; Volume 10.
23. Cambria, E.; Havasi, C.; Hussain, A. Senticnet 2: A semantic and affective resource for opinion mining and sentiment analysis. In Proceedings of the Twenty-Fifth International FLAIRS Conference, Marco Island, FL, USA, 23–25 May 2012.
24. Cambria, E.; Olsher, D.; Rajagopal, D. SenticNet 3: A common and common-sense knowledge base for cognition-driven sentiment analysis. In Proceedings of the AAAI Conference on Artificial Intelligence, Quebec City, QC, Canada, 27–31 July 2014; Volume 28.
25. Cambria, E.; Poria, S.; Bajpai, R.; Schuller, B. SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives. In *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, Osaka, Japan, 11–16 December 2016*; The COLING 2016 Organizing Committee: Osaka, Japan, 2016.

26. Cambria, E.; Poria, S.; Hazarika, D.; Kwok, K. SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.
27. Cambria, E.; Li, Y.; Xing, F.Z.; Poria, S.; Kwok, K. SenticNet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, Online , 19–23 October 2020; pp. 105–114.
28. Liang, B.; Su, H.; Gui, L.; Cambria, E.; Xu, R. Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks. *Knowl.-Based Syst.* **2022**, *235*, 107643. [CrossRef]
29. Pennington, J.; Socher, R.; Manning, C.D. Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1532–1543.
30. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
31. Ma, Y.; Peng, H.; Cambria, E. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018, Volume 32.
32. Song, Y.; Wang, J.; Jiang, T.; Liu, Z.; Rao, Y. Attentional encoder network for targeted sentiment classification. *arXiv* **2019**, arXiv:1902.09314.
33. Manandhar, S. Semeval-2014 task 4: Aspect based sentiment analysis. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland, 23–24 August 2014; pp. 27–35.
34. Pontiki, M.; Galanis, D.; Papageorgiou, H.; Manandhar, S.; Androutsopoulos, I. Semeval-2015 task 12: Aspect based sentiment analysis. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, CO, USA, 4–5 June 2015; pp. 486–495.
35. Pontiki, M.; Galanis, D.; Papageorgiou, H.; Androutsopoulos, I.; Manandhar, S.; AL-Smadi, M.; Al-Ayyoub, M.; Zhao, Y.; Qin, B.; De Clercq, O.; et al. Semeval-2016 task 5: Aspect based sentiment analysis. In Proceedings of the ProWorkshop on Semantic Evaluation (SemEval-2016), Association for Computational Linguistics, San Diego, CA, USA, 16–17 June 2016; pp. 19–30.
36. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
37. Kiritchenko, S.; Zhu, X.; Cherry, C.; Mohammad, S. NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland, 23–24 August 2014; pp. 437–442.
38. Wang, K.; Shen, W.; Yang, Y.; Quan, X.; Wang, R. Relational graph attention network for aspect-based sentiment analysis. *arXiv* **2020**, arXiv:2004.12362.
39. Gao, Z.; Feng, A.; Song, X.; Wu, X. Target-dependent sentiment classification with BERT. *IEEE Access* **2019**, *7*, 154290–154299. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.